AD-A113 953   TEXAS CHRISTIAN UNIV FORT WORTH INST OF BEHAVIORAL R--ETC F/G 5/10
ORGANIZATIONAL CLIMATE: ANOTHER LOOK AT A POTENTIALLY IMPORTANT--ETC(U)
APR 82 L R JAMES   N00014-80-C-0315
UNCLASSIFIED IBR-82-4   NL

1 OF 1
AD A
113953

END
DATE
FILMED
05-82
DTIC

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: IBR 82-4	2. GOVT ACCESSION NO.: AD-A113 953
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Organizational Climate: Another Look at a Potentially Important Construct	5. TYPE OF REPORT & PERIOD COVERED: Technical Report
	6. PERFORMING ORG. REPORT NUMBER: IBR 82-4
7. AUTHOR(s): Lawrence R. James	8. CONTRACT OR GRANT NUMBER(s): N00014-80-C-0315
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Institute of Behavioral Research, Texas Christian University, Fort Worth, Texas 76129	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR170-904
11. CONTROLLING OFFICE NAME AND ADDRESS: Organizational Effectiveness Research Programs, Office of Naval Research (Code 452), Arlington, Virginia 22217	12. REPORT DATE: April 7, 1982
	13. NUMBER OF PAGES: 33
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

DTIC ELECTE APR 28 1982 S D

D

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Agreement
Interrater Reliability
Intraclass Correlation
Organizational Climate

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Organizational climate is examined from the standpoint of the need to demonstrate interrater reliability (agreement) among individuals' perceptions of psychological climate. It is shown that if prior estimates of interrater reliability are accurate, then organizational climate is a moot issue. It is then demonstrated that at least some prior estimating procedures likely provided underestimates of interrater/perceptual agreement. A new procedure for estimating interrater ___

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD A113953

DTIC FILE COPY

82 04 28 050

9. School of Psychology
Georgia Institute of Technology
Atlanta, Georgia 30332

20. reliability is suggested and illustrated. The new procedure provided substantially higher estimates of interrater reliability than prior methods. It is concluded that the concept of organizational climate is potentially salvageable, although more attention needs to be given to the appropriate level of explanation for climate variables.

Organizational Climate:

Another Look at a Potentially Important Construct

In a recent discussion of measurement models in climate research, James (1982) recommended that a decision of whether to aggregate individuals' climate scores should be a function of the magnitude of an intraclass correlation estimate of interrater reliability. This recommendation was based on the following rationale: (a) the basic unit of theory (unit of analysis) for climate is individuals' perceptions of their psychological climate (James & Sells, 1981; Jones & James, 1979; Joyce & Slocum, 1979; Schneider, in press); (b) a composition theory relating psychological climate scores to aggregate psychological climate scores (e.g., organizational climate scores) may be established if the perceptions of psychological climate are shared among the individuals on whom the aggregate is computed (Roberts, Hulin, & Rousseau, 1978); and (c) the typical design employed in climate studies is a random effects, one-way analysis of variance (ANOVA), from which, given reasonable satisfaction of assumptions, it is possible to estimate interrater reliability (perceptual agreement, degree perceptions are shared) by the intraclass correlation equation for the reliability of a single rating or measurement [referred to here as ICC(1)--cf. Bartko, 1976; Ebel, 1951; Shrout & Fleiss, 1979; Winer, 1971].

The objectives of the present paper represent, in part, a continuation of the discussion above. It is suggested that the criterion for perceptual agreement and aggregation of psychological climate scores is a reasonably high ICC(1). Based on this criterion, it is shown that legitimate indices of interrater reliability render organizational climate a moot issue, where the term organizational climate is used to refer to a field of research which involves any type of aggregate psychological climate scores (Jones & James, 1979; Schneider, in press). It is then demonstrated that estimates of interrater

reliability based on an ICC(1) approach may, under specified conditions, furnish *serious underestimates* of interrater reliability. Finally, a new method for estimating interrater reliability is overviewed, and an empirical illustration is used to show that it is possible to achieve high levels of interrater reliability on climate data. The conclusion reached is that organizational climate is a salvageable construct.

A Criterion for Perceptual Agreement in Climate Research

Climate has been reviewed extensively in recent years, the output focused mainly on restatements of prior positions, reviews of these positions, and reviews of reviews (Campbell, Dunnette, Lawler, & Weick, 1970; Hellriegel & Slocum, 1974; Insel & Moos, 1974; James, Hater, Gent, & Bruni, 1978; James & Jones, 1974; James & Sells, 1981; Jones & James, 1979; Joyce & Slocum, 1979; Naylor, Pritchard, & Ilgen, 1980; Payne, Fineman, & Wall, 1976; Payne & Pugh, 1976; Powell & Butterfield, 1978; Schneider, 1975, in press; Schneider, Parkington, & Buxton, 1980; Woodman & King, 1978). Represented ubiquitously in these reviews is the logic that perceptual agreement should precede aggregation of climate scores. Yet, a criterion for an acceptable level of perceptual agreement--that is, a level that justifies aggregation--remains obscure (James, 1982). Exceptions to this rule include Guion (1973), who recommended that agreement indices should not depart significiantly from 1.00, and Roberts et al. (1978) and Schneider (in press), who recommended that within-organization variance in climate perceptions should be small in relation to between-organization variance. The fact that these two recommendations are statistically miles apart is easily demonstrated if we apply their statistical implications to the typical experimental design used in climate research.

Suppose that we have n_k individuals nested in each of K (k=1,...,K) organizations. For the present, it is presumed that the assumptions for a one-way

random effects ANOVA and computation of an ICC(1) have been reasonably satisfied (e.g., randomly selected organizations and individuals, homogeneity of variance). The ANOVA employs the K organizations as treatments and the n_k scores on a climate variable in each organization as values on the dependent variable. The "empirical criterion" for agreement for Roberts et al. (1978) and Schneider (in press) appears to be a significant F ratio, which connotes significantly greater between-organization variance than within-organization variance. Note that no point estimate of interrater reliability is required. This suggests, for example, that with large samples an ICC(1) of .05 is acceptable as long as the F ratio is significant [cf. Jones & James, 1979 for point estimates of ICC(1) with large samples]. The Guion (1973) criterion is much more stringent. It implies, for example, that in each of the K organizations the variance on the climate variable should not depart significantly from zero (see Li, 1964 for a chi-square test for variances). This suggests that not only should the F ratio be significant, but also that the ICC(1) should approach 1.00.

The position advocated in this paper is that a criterion for perceptual agreement requires first a point estimate of interrater reliability. To demonstrate merely that an F ratio is significant is of trivial concern in relation to the magnitude of the interrater reliability estimate, especially when $N (N = \Sigma n_k)$ is large (Cohen, 1960). Thus, the Roberts et al. (1978) and Schneider (in press) criterion is not regarded as sufficient for justifying aggregation of climate scores. The implied necessity for a point estimate of interrater reliability approaching 1.00 (Guion, 1973) is regarded as too stringent. Consider, for example, that the conventional criterion for computing an aggregate over items--that is, a composite score for each individual on items designed to assess the same construct--is an internal consistency reliability (e.g., coefficient α) of

.70 and above (in exploratory studies). If we were to require $\underline{\alpha}$'s approaching 1.00, few item composites would be computed. I submit that the same conventional criterion can be used as a criterion for interrater reliability and aggregation of climate scores over individuals. Specifically, given the design in question, it is recommended that an ICC(1) of .70 should be employed as a lower-bound criterion for justifying aggregation of climate scores over individuals.

Is Organizational Climate a Moot Issue?

James (1982) summarized estimates of perceptual agreement in climate studies and reported that the range of estimates varied from .00 to .50, with a median of approximately .12. The estimates included in the summary were based on either ICC(1) or estimates of the proportion of variance in individuals' perceptions associated with variation among environments (eta-squares, omega-squares). For reasons explained in that article, estimates based on aggregates were considered biased and excluded from the summary. Also excluded were estimates of interrater reliability based on correlations among profiles (e.g., a correlation between two raters' scores on a set of climate dimensions) for reasons discussed by numerous authors (cf. Cronbach & Gleser, 1953), and a study by Howe (1977), which confounded stability of perceptions over time with agreement among perceptions at a particular point in time.

Given that legitimate estimates of interrater reliability do not exceed .50, it follows if we were to adopt a point estimate of interrater reliability equal to or greater than .70 as the operational criterion for perceptual agreement and aggregation, then organizational climate as presently conceived is a moot issue. Or is it?

Appropriateness of the Intraclass Correlation in Climate Studies

The objective of this section is to suggest that the intraclass correlation, and other statistics that employ a between-group versus within-group form of design (eta-square, omega-square), may have provided substantial underestimates

of interrater reliability in at least some prior climate studies. Discussion focuses on *ICC*(1), and is based on a recent statistical paper by James, Wolf, and Demaree (Note 1).

Associated with the *ICC*(1) statistic are a number of assumptions underlying the ANOVA procedure on which it is based. One assumption is that the environments employed in a study comprise a random sample of environments from a *heterogeneous* population of environments. The somewhat subtle implication of this assumption is that if a mean (aggregate) climate score is computed over the n_k individuals in each environment, then these means will vary among environments, especially in a condition of high interrater reliability. To be specific, between-environment variance in mean climate perceptions is a prerequisite for high interrater reliability. Now, consider the statistical facts that if (a) little variation exists among the K *mean* climate perceptions for K environments, and if (b) perceivers in each of the K environments agree almost perfectly (i.e., within-environment variance is close to zero), then (c) the *ICC*(1) estimate of interrater reliability will be low. Note that we have a condition of almost perfect agreement within environments and an estimate of interrater reliability that is conceivably zero (or even negative in value). In effect, this is the restriction of range problem extended to estimates of interrater reliability, where by restriction of range is meant little or no variation among the mean climate perceptions over environments. (The same logic applies to eta-square and omega-square, although these statistics may themselves differ; cf. Maxwell, Camp, & Arvey, 1981).

These points are easily illustrated statistically. The data presented in Table 1 consist of hypothetical scores on a climate item X which has five discrete, approximately equally spaced alternatives (e.g., a Likert scale--cf. Cooper, 1976; Hsu, 1979). Frequencies of responses are shown for 20 different

individuals in each of two environments. The frequencies of responses indicate that the individuals in each environment tend to agree, which is reflected by the small within-environment variances (.211 and .261). However, *ICC*(1) is -.047, which is regarded as .00 (Bartko, 1976). This low *ICC* is clearly an underestimate of true agreement, and may be attributed directly to the essential absence of variation among the aggregate climate scores (3.00 and 3.05).

Insert Table 1 about here

Data such as presented in Table 1 stimulate the following question: *Why should the level of agreement within an environment be contingent on differences among environments*? That is, in its most direct form, interrater reliability and agreement address the question of whether people in a particular environment, or people in each one of a set of environments, agree with respect to their perceptions. *This question neither assumes nor requires that differences exist among environments*. Of course, if environments were sampled randomly from a heterogeneous population of environments in which mean climate perceptions were expected to vary, then we would not anticipate a restriction of range problem such as illustrated in Table 1. This point is discussed below. It is also noteworthy that if the level of agreement varies as a function of environment (i.e., the level of agreement is *not* the same or similar across environments), then the ANOVA-based *ICC*(1) approach cannot be used because the homogeneity of variance assumption is violated. Thus, even if one wanted to include between-environment differences in an interrater reliability estimate, one could not do so legitimately.

In sum, use of a statistic such as the *ICC*(1) that relies on between-environment differences will result in an underestimate of interrater reliability (agreement) if the following conditions exist: (a) mean climate scores do not

vary meaningfully among environments, and (b) individuals within environments tend to agree.

A case can be built that these two conditions apply to at least some climate studies. The case for low variation among mean climate scores over a set of $\underline{K}$ environments is predicated on the fact that many climate studies have employed samples of environments from the same basic system or, more typically, subsystem type. It is not uncommon to find the sample in a particular study limited to banks, to classrooms, to dormitories, to hospitals, to life insurance agencies, or to divisions aboard Navy ships. Now consider the possibility that variation among mean climate scores is likely to be restricted if all environments in a sample are of the same or similar basic type, regardless of whether the environments were randomly sampled from within this type. That is, sampling of environments from a homogeneous type of environment, in relation to a more heterogeneous population of environmental types, is likely to lead to restricted variances on situational attributes believed to be causes of climate perceptions, such as technology, structure, systems norms and values, and processes (e.g., communication, leadership, and rewards). It follows that if (a) individuals' climate perceptions are a (partial) function of situational attributes, and if (b) sampling from a homogeneous environmental type results in restricted ranges on situational attributes, then (c) the range should also be restricted on individuals' perceptions, and, therefore, means of individuals' perceptions.

The case for low within-environment variation among individuals' perceptions is based in part on the argument above and in part on a recent report by Schneider (in press). Range restriction in regard to the type of environment studied suggests similarity of perceptions because of similarity in situational stimuli. However, similarity of stimuli is not sufficient to guarantee low within-environment variation in perceptions. My colleagues and I have argued on a number of occasions that individuals with different cognitive construction

competencies, encoding abilities, self-regulatory systems, beliefs, needs, values, and self-concepts will be predisposed to differ in what they perceive as ambiguous, challenging, fair, friendly, supportive, and so forth (cf. James, Hater, Gent, & Bruni, 1978; James & Sells, 1981). That is, psychological climates associated with the same or similar actual environments are likely to differ for <u>different types of individuals</u>, and the reasons for these differences are not only psychologically important, but also they can be reliably measured and related to climate perceptions (cf. James, Gent, Hater, & Coray, 1979; James, Hater, & Jones, 1981; James & Jones, 1980).

On the other hand, if, as Schneider (in press) suggests, the environments in question are composed of <u>similar types of individuals</u>, then I agree with his conclusion that the likelihood of variation in perceptions due to individual differences is reduced. If placement in a particular job, office, position, or role is subject to rigorous selection standards that relate, directly or indirectly, to cognitive information processing competencies and predispositions (e.g., achievement motivation, cognitive complexity, intelligence, perceived competence, and self-esteem), then the resulting relative similarity among individuals suggests a <u>relative</u> similarity among perceptions of climate. Of perhaps equal importance is the degree to which individuals with relative similarities in attributes not necessarily related to formal selection processes (e.g., cosmopolitan vs. local orientation, expectancies, locus of control, need for affiliation) are attracted to (self-select) a particular job, office, position, or role. Here again, relative similarity in individual attributes suggests relative similarity in perceptions. Furthermore, relative similarity among individuals resulting from formal and/or self-selection processes generates forces toward perceptual agreement because (a) environments tend to be shaped to fit the type of individuals who select, and are selected, to work in them, which implies similarity of within-environmental stimuli for similar

types of individuals (cf. Endler & Magnusson, 1976; James et al., 1978); and (b) the meaning imputed to an environment by an individual is more likely to be socially influenced by other individuals in that environment if the perceiver views the others as similar to himself/herself than if the others are viewed as different (cf. Stotland & Canon, 1972).

In summary, the following two situations appear to be conducive to under-estimation of interrater reliability/agreement when estimation is based on an ANOVA design and the ICC(1).

1) Sampling of homogeneous environments, which implies restriction of range in the types of situational stimuli perceived in each environment and a similarity of stimuli across environments.

2) Similar types of individuals within homogeneous environments, resulting from rigorous formal selection processes and/or self-selection processes. Similarity among individuals implies, in a relative sense, a narrow range of individual differences in cognitive information processing competencies and predispositions. This in turn suggests relative similarities in the psycholog-ical meaning and significance imputed to environments (i.e., similar psychological climates). It also suggests similar shaping of environmental stimuli and social influence processes.

These two situations are conducive to the statistical conditions of low within-environment variation resulting from similar types of individuals perceiv-ing similar types of stimuli, and low between-environment variation in mean (aggregate) climate scores. An alternative to the ICC(1) approach is indicated for estimating interrater reliability/agreement if these two statistical conditions are operative, or perhaps even partially operative. Such an alternative was proposed by James et al. (Note 1), and is reviewed below.

An Overview of a New Method for Estimating Interrater Reliability in Climate Studies

The proposed procedure was based on prior work by Finn (1970) and Cooper (1976), and employs a within-group design in which interrater reliability is estimated separately for each group (i.e., environment). For each group, interrater reliability is defined as the degree to which raters (perceivers) agree with respect to their ratings (perceptions) of a particular target (e.g., the organization) on a particular rating (climate) scale (e.g., the equity of an organization's pay and benefit system). A within-group design is used because we desire an estimate of interrater reliability for each group that is not a function of between-group variation. Thus, the estimate will not be affected by lack of variation in group means. Furthermore, lack of homogeneity of within-group variation is not a concern inasmuch as a separate estimate of reliability is computed for each group. Consequently, agreement may vary as a function of environment and we may still estimate agreement for each group.

The proposed procedure views interrater reliability (agreement) within a group as a function of two variances, namely (a) the observed variance among the ratings on a climate item X, designated s_X^2, and (b) the expected variance among the ratings on climate item X in a condition of no agreement, designated σ_E^2. An $s_X^2 = 0$ indicates perfect agreement; however, s_X^2 is not usually equal to zero and thus we must ascertain the degree to which raters in a group agreed. This is accomplished by comparing s_X^2 to σ_E^2, where σ_E^2 is the variance on item X that would be expected if raters responded randomly, which implies zero interrater reliability and no agreement (cf. Finn, 1970). Thus, σ_E^2 functions as a statistical benchmark for random responsing and absence of agreement. It follows that (a) the value of the proportion indicated by s_X^2/σ_E^2 reflects the amount of random error variance in the observed ratings, and (b) $1 - (s_X^2/\sigma_E^2)$ is a reliability coefficient because it indicates the proportion of nonerror

variance in the observed ratings (Finn, 1970; James et al., Note 1).

It is important to note that $\underline{\sigma_E}^2$ is a statistical abstraction. Whether raters in a particular group would ever respond in a sheerly random fashion is irrelevant to the use of hypothetical random responding as a statistical benchmark for assessing the extent to which the variance of a set of actual responses, indicated by $\underline{s_X}^2$, resembles the expected variance of a set of random responses, indicated by $\underline{\sigma_E}^2$. The assumption of random responding also provides a simple method for computing $\underline{\sigma_E}^2$. Random responding implies that each alternative on the rating scale for item $\underline{X}$ has an equal likelihood of response. This in turn implies that the distribution of scores over alternatives is rectangular or uniform. Consequently, $\underline{\sigma_E}^2$ may be calculated using the equation for the variance of a discrete, uniform distribution. This equation is: $\underline{\sigma_E}^2 = (\underline{A}^2-1)/12$, where $\underline{A}$ corresponds to the number of discrete alternatives on item $\underline{X}$. $\underline{\sigma_E}^2$ is a population parameter, and thus sample size does not enter into its calculation.

In summary, building on prior work by Finn (1970) and Cooper (1976), James et al. (Note 1) derived the following equation for estimating interrater reliability for a single group of individuals on a single item.

$$\underline{r_{WG}} = 1 - (\underline{s_X}^2/\underline{\sigma_E}^2) \qquad (1)$$

where:

$\underline{r_{WG}}$ = within-group interrater reliability for a single group of raters on a single item $\underline{X}$.

$\underline{s_X}^2$ = the observed variance on item $\underline{X}$ in the group. Assumptions associated with $\underline{s_X}^2$ (and $\underline{\sigma_E}^2$) are that raters responded independently (this does not preclude prior social influence processes), and that $\underline{X}$ is a discrete random variable with multiple alternatives arranged on an approximately interval

scale (such as a Likert item--cf. Cooper, 1976).

σ_E^2 = the variance on item X that would be expected if the raters responded randomly, which implies zero interrater reliability and no agreement. σ_E^2 is calculated by the equation $(A^2-1)/12$, where A is the number of alternatives on item X (the scale on X is $1,\ldots,A$).

Equation 1 is easily interpreted. If $s_X^2 = 0$, then $r_{WG} = 1.0$; that is, no variance on X results in perfect interrater reliability (agreement). Conversely, if raters were to respond randomly, then $s_X^2 \cong \sigma_E^2$, and $r_{WG} \cong 0$. The typical situation in research is $0 < s_X^2 < \sigma_E^2$. Equation 1 indicates that as s_X^2 approaches σ_E^2, interrater reliability decreases, and as s_X^2 becomes progressively smaller than σ_E^2, interrater reliability increases.

The use of Eq. 1 is illustrated by application to the data in Table 1. Item X has five alternatives, or $A = 5$, and thus σ_E^2 is equal to $(5^2 - 1)/12$, or 2.0, in each of the two groups. The observed variance (s_X^2) on X in Group 1 is .211, and the estimate of r_{WG} provided by Eq. 1 is .89 [i.e., 1 - (.211/2.0)]. Using similar procedures, the estimate of r_{WG} in Group 2 is 1 - (.261/2.0), or .87. Given the similarity of these two estimates (and the observed variances), the estimates were averaged to furnish a value of .88. The value of .88 is obviously different than the ICC(1) of .00, and it is equally obvious that each r_{WG} and the average r_{WG} are more consistent with the data than the ICC(1).

It should be noted that averaging the separate r_{WG} across groups is not recommended if the r_{WG} are dissimilar. A homogeneity of variance test on observed variances (i.e., the s_X^2s) assists in ascertaining whether to average r_{WG}s in nonobvious situations.

Interrater reliability for composite scores. Data employed in climate studies are often based on a composite score rather than a single item. For example, each member of a workgroup rates that group on a set of items designed to measure workgroup cooperativeness. A composite score is then calculated for each rater by computing a sum or a mean over the items, and it is these scores that are entered into the within-group interrater reliability (agreement) analysis. Based on rationale by Finn (1970), James et al. (Note 1) derived an equation for estimating the interrater reliability among raters' composite scores on a set of J ($j=1,\ldots,J$) items in a single group. The derivation was based on the assumptions that (a) the J items represent a random sample of items from a single, well-defined domain of items (cf. Lord & Novick, 1968); (b) the raters in each group are randomly sampled from a population of raters to which inferences will be made (which allows the population of raters to be homogeneous), and (c) the item variances and interitem covariances are equal, respectively, in the rater population, which implies that the items are "essentially parallel" indicators of the same construct.

An example of the design in question is presented in Section A of Table 2. The data represent ratings (i.e., item responses) provided by six raters ($i=1,\ldots,n_k=1,\ldots,6$) on four essentially parallel items that measure the same climate dimension. Each of the four items ($J=4$) employs the same seven discrete, approximately equally spaced alternatives ($A=7$).

Insert Table 2 about here

The generally recommended statistical procedure for estimating interrater reliability for multiple ratings in a within-group design should not be used here. As shown in Section B of Table 2, the within-group ICC is approximately .00 [equation for ICC (2,1), Shrout & Fleiss, 1979]. This is due to the fact

that the items have essentially identical means, from which it follows that the between-item mean square is close to zero. The within-group *ICC* can only assume high values when between-item variance is larger than within-item variance. Given essentially parallel items, this is not likely to be the case, and the within-group *ICC* underestimates interrater reliability.

The procedure described by James et al. (Note 1) is designed to estimate interrater reliability among rater composite scores in the form of means, designated $\bar{X}_i$. The $\bar{X}_i$ are displayed at the bottom of the data matrix in Section A of Table 2. The estimating equation takes a number of forms; the most direct for computing purposes is as follows:

$$r_{WG(\bar{X}_i)} = \frac{J[1 - (\overline{s^2_{X_j}}/\sigma_E^2)]}{J[1 - (\overline{s^2_{X_j}}/\sigma_E^2)] + (\overline{s^2_{X_j}}/\sigma_E^2)} \tag{2}$$

where:

$r_{WG(\bar{X}_i)}$ = within-group interrater reliability for mean rater scores (the $\bar{X}_i$) on J essentially parallel items.

$\overline{s_{X_j}^2}$ = the mean of the observed variances on the J items--it is assumed here that each of the J items employs the same seven alternatives.

σ_E^2 = same definition as before, namely the expected variance of an item in a condition of zero interrater reliability and no agreement. Technically, the mean σ_E^2, or $\overline{\sigma_E^2}$, should be used in Eq. 2, but with A=7 for all items, $\overline{\sigma_E^2} = \sigma_E^2$.

The use of Eq. 2 is illustrated in Section C of Table 2. The estimate of $r_{WG(\bar{X}_i)}$ is .98, which contrasts sharply with the within-group *ICC* of .00. It

is also clear that an interrater reliability of .98 is a more accurate reflection of the data than a .00. To be fair here, one could argue that the within-group ICC is low because the items were not sampled randomly from a heterogeneous population of items, thus violating the implicit ANOVA assumption of variation among items means. The within-group ICC was included only to demonstrate its inapplicability.

Equation 2, like Eq. 1, may be applied in each of K groups, and the resulting $r_{WG(\bar{X}_i)}$ may be averaged over the K groups if the separate $r_{WG(\bar{X}_i)}$ are similar. Homogeneity of variance tests on the mean item variances over the K groups might be employed to help to decide whether to average the $r_{WG(\bar{X}_i)}$. Finally, it is suggested that if the decision is to average, and the n_k differ, there would be little reason to weight the $r_{WG(\bar{X}_i)}$ by n_k because the $r_{WG(\bar{X}_i)}$ should be similar. This applies also to averaging r_{WG}.

In summary, the discussions above summarize the use of r_{WG} and $r_{WG(\bar{X}_i)}$. Statistical derivations and discussions of potential criticisms of the procedures are presented in James et al. (Note 1). It is noted here that very small n_k (e.g., less than 10 individuals in a group) may lead to unstable results, and that very short (e.g., $A \leq 3$) or very long (e.g., $A > 9$) item scales may produce artifical results. Additional points developed more fully in the James et al. paper are (a) although the theoretical distribution on an item X may be normal (Hsu, 1979; Selvage, 1976), a rectangular (uniform) distribution should be used to calculate σ_E^2 because the rectangular distribution, and not the normal distribution, models random responses (the normal distribution models partial agreement because of central tendency); (b) the calculation of σ_E^2 may be based on an assumed underlying continuous, rather than discrete, distribution by

using $(\underline{A}-1)^2/12$ to calculate σ_E^2 (Selvage, 1976); (c) like ICC(1), the estimates of r_{WG} and $r_{WG(\bar{X}_i)}$ are biased, but the bias is expected to be minimal for small n_k and essentially neglible for large n_k; and (d) also like ICC(1), r_{WG} and $r_{WG(\bar{X}_i)}$ can assume values of less than .00, in which case the value is set equal to .00 because all observed distributions on an item X that result in negative values are due to serious degrees of disagreement [the same recommendation was made for ICC(1) by Bartko, 1976].

Empirical Comparison of Between-Group and Within-Group Approaches

The data employed in this illustration were collected by David W. Bracken as part of a dissertation project at the Georgia Institute of Technology, and loaned to the present investigator to demonstrate statistical procedures. The data met the two situations and statistical conditions discussed earlier in which an ICC(1) procedure would be expected to provide an underestimate of interrater reliability. Statistical conditions are discussed shortly. Of initial concern is that Situation 1 was satisfied inasmuch as all environments were of the same organization subtype. The environmental sample consisted of field offices of a large business machines company. Each office (a) operated as a self-contained subsystem; (b) performed the same functions, namely marketing, installing, and servicing small business machines; and (c) had the same hierarchical/functional differentiation, where the staff consisted of managerial personnel, marketing personnel, supervisors, technicians (see below), and clerical personnel. All offices were located in the United States, with the exception of one location in Puerto Rico. The offices varied in size, but size was not related to the data of interest here.

Situation 2 refers to relative homogeneity of within-office variance on

individual difference variables that could influence scores on climate variables. This situation was partially satisfied in the following manner. The parent corporation supported a study designed to ascertain if climate moderated relationships between scores on selection tests and performance. The study focused exclusively on the position of technician, which is similarly described for all offices as installing and servicing business machines. For the present study, the relative homogeneity of variance on individual difference variables for technicians was demonstrated by comparing selection test data to published norms in test manuals, where the most heterogeneous norm samples--high school students--were selected for comparison purposes. The tests included the Bennett Test of Mechanical Comprehension and the Gordon Personal Inventory and Personal Profile. The personality data were of primary interest here because personality comprises a salient basis for predispositions toward assignment of meaning in higher-order cognitive processing (cf. James & Jones, 1980; James et al., 1978; James et al., 1979; Jones & Gerard, 1967; Kim, 1980; Mischel, 1973; Stotland & Canon, 1972).

Test data were available on approximately 2,800 technicians; all offices (K=87) were represented in this sample. These data were based on test results obtained since 1975, the time at which the parent corporation initiated a formal reporting program. Company personnel regarded these data as representative of most technicians because the tests had been used in the same fashion for a number of years prior to the initiation of the reporting program. While it was not possible to confirm/disconfirm this assumption empirically, the results for interrater reliability on climate perceptions reported shortly suggest that the assumption is valid.

The relative homogeneity of variance on individual difference variables is indicated by the statistics reported in Table 3. These results demonstrate that (a) the means on test scores for the technician sample were, with one

exception (sociability), far above the averages of the norm samples, as indicated by percentile ranks ranging from 67 to 95; and (b) on the average, the variances of test scores on the technician sample were approximately one-half as large as those on the norm samples. Furthermore, as shown in column three of Table 3, no meaningful variation in technicians' test scores was associated with differences among offices, which is indicated by the small eta-squares (eta-squares were based on one-way ANOVAs using office as the independent variable). These results suggest that variation on individual difference variables was relatively restricted for the technician sample and that technicians were, on the average, similar across offices. This does not imply that all technicians were the same or reported the same climate perceptions; for example, sufficient variation remained to conduct a principal components analysis on the climate scores for the technician sample.

Insert Table 3 about here

The sample of technicians employed in the interrater reliability analyses on climate perceptions consisted of 7,180 individuals from 60 offices. (These data were collected in the first phase of the study; data collected in later phases were essentially identical to those reported here.) The technicians completed a 102 item climate questionnaire, developed specifically for technicians, in 1980. Principal components analysis furnished 13 components that were interpretable as cognitive representations of the work environment and had scale reliabilities of .70 or greater (coefficient alpha). Abbreviated designations of 11 of the climate dimensions for which "office" was a potentially appropriate level of explanation are presented in Table 4. The remaining two climate dimensions (supervisor support, workgroup cooperativeness) are not included because different levels of explanation were indicated (e.g., different

supervisors and workgroups within a particular office). Also shown in Table 4 is the number of items per climate dimension, dimension internal consistency estimates (coefficient alpha), estimates of interrater reliability furnished by an ICC(1) approach, the variance of the mean climate scores for the 60 offices, the average within-office variance on each climate dimension, and estimates of interrater reliability supplied by $r_{WG(\bar{X}_i)}$, which are reported in terms of the range of estimates for the 60 offices and the percent of offices for which the estimate was .70 or above.

Insert Table 4 about here

The intraclass correlations were computed by first calculating composite (mean) scores on each of the 11 climate dimensions for each technician. These scores were based on a mean of the scores on the items that loaded on each climate component.[1] For each climate dimension, a random effects, one-way ANOVA was conducted to obtain estimates of the between-office and within-office mean squares. The ICC(1) equation was then employed to estimate interrater reliability, using a harmonic mean of 75.05 for office size (office size ranged from 5 to 215). As shown in the ICC(1) column in Table 4, the estimates of interrater reliability were uniformly low, ranging from .01 to .10. These results are generally consistent with prior climate studies, albeit the ICC(1)s are on the low side.

An explanation for these low ICC(1)s is furnished in columns 4 and 5 of Table 4. A mean (aggregate) of the technicians' climate scores was computed for each climate dimension for each office, thus furnishing 60 mean office climate scores for each climate dimension. The variance of the means for each climate dimension on the sample of 60 offices is reported in Column 4. The range of

possible mean office scores is 1 to 5. The range of the 11 variances is .01 to .07, which on a five-point scale suggests restriction of range on the mean climate scores for offices. These results support the prior contention of homogeneity of the environmental sample. The average within-office variances on technicians' climate composite (mean) scores are presented in column 5. The values reflect the variation of the climate composite scores in an office about the mean for that office, averaged over the 60 offices. Again using a scale of 1 to 5, these variances generally reflect low within-office variation on the climate composite scores (exceptions occurred for dimensions 2 and 11).

The data shown in columns 4 and 5 of Table 4 are consistent with the two statistical conditions (low variation among office means and low within-office variation in technicians' climate scores) that suggest a low ICC(1). Low within-office variation implies further that the interrater reliabilities based on $r_{WG(\bar{X}_i)}$ should be substantially higher than those based on ICC(1). The estimates of $r_{WG(\bar{X}_i)}$ were based on applications of Eq. 2 in each of the 60 offices for each of the climate dimensions. The $r_{WG(\bar{X}_i)}$s for each climate dimension differed somewhat across the 60 offices, and thus ranges and the percent of estimates greater than or equal to .70 are reported in columns 6 and 7 of Table 4. The estimates furnished by this analysis were substantially higher than the estimates provided by the ICC(1) procedure. For example, at least 88% of the 60 offices had $r_{WG(\bar{X}_i)}$s of .70 or greater on 9 of the 11 climate dimensions. Moreover, even for the remaining two dimensions (dimensions 2 and 11), not only did some offices (22% and 65%, respectively) have $r_{WG(\bar{X}_i)}$s $\geq$ .70, but also the lowest value in the range of $r_{WG(\bar{X}_i)}$s exceeded ICC(1). It should also be mentioned that the values

of $\underline{r_{WG(\overline{X}_i)}}$ and the percent of $\underline{r_{WG(\overline{X}_i)}}$s $\geq$ .70 tended to be larger for the climate dimensions with the larger number of items (dimensions 1, 3, and 10). This is a result of the fact that $\underline{r_{WG(\overline{X}_i)}}$ is a function of $\underline{J}$, the number of items in a composite (see Eq. 2). A substantive interpretation of this function is that the mean score per rater (i.e., $\underline{\overline{X}_i}$) will contain less random measurement error as the number of essentially parallel items in a composite increases. Thus, the estimate of interrater reliability among composite scores will be less influenced by random error in the original item measurements as the number of essentially parallel items in the composite increases. On the other hand, inspection of Eq. 2 shows that a large $\underline{J}$ does not guarantee a large $\underline{r_{WG(\overline{X}_i)}}$ (e.g., if $\underline{\overline{s^2_{X_j}}} \sim \underline{\sigma_E}^2$, the $\underline{r_{WG(\overline{X}_i)}} \cong 0$).

In summary, the results for the $\underline{r_{WG(\overline{X}_i)}}$ analysis, in contrast to the results for the ICC(1) analysis, suggest that the technicians in most offices tended to agree with respect to their climate perceptions on 9 out of a possible 11 climate dimensions. Consequently, mean or aggregate climate scores for technicians could be calculated for 9 climate dimensions in almost all offices, and used to describe the shared psychological environment (climate) among technicians in that office.

Conclusions

A conclusion that is not warranted by the discussion and illustration is that all prior climate studies that employed between-group designs reported underestimates of interrater reliability (perceptual agreement). For example, given homogeneity of within-group variance and moderate to large between-group differences in group mean climate scores, the ICC(1) can assume high and reasonably accurate estimates of interrater reliability (James et al., Note 1). The primary problem occurs when between-group differences are small and within-group variation is low, which is most likely to occur when groups are sampled from the same environmental type or subtype and the range on individual difference

variables is restricted due to formal and self-selection procedures. Inasmuch as many climate studies employ at least samples of environments from the same environmental type, the potential for underestimates of interrater reliability is present. Unfortunately, published studies do not furnish sufficient data for reanalysis using the methods described here.

It is strongly recommended that climate researchers reexamine their data in light of the substantive and statistical points made in this paper. I believe that it is reasonable to assume that such reexamination will lead to different conclusions for at least some studies. That is, if the empirical illustration reported here is generalizable, then it is quite possible that estimates of interrater reliability will be higher, perhaps much higher, than those reported previously. On the other hand, the empirical illustration may be idiosyncratic and not generalizable. One could argue that the estimates of interrater reliability were higher than would generally be expected because only individuals from the same position (i.e., technician) were included in the analysis. I will attempt to counter this argument by suggesting that individuals in different positions in an organization are likely (a) to experience different situational stimuli, which contributes to different perceptions of climate (cf. Newman, 1975), and (b) to vary in regard to individual variables which affect the meaning assigned to situational stimuli. The latter concern is viewed as a function of the formal selection and self-selection processes discussed earlier, and as a function of experience in the organization (e.g., increases in self-esteem resulting from promotions). This suggests that we should consider position, or perhaps families of similar positions, as a key variable on which to base agreement analyses. This, of course, is an empirical question that can be addressed in future research.

In conclusion, it is submitted that estimates of interrater reliability equal to or above .70 should not be all that uncommon in climate research if

(a) the data have satisfactory psychometric properties (e.g., high scale reliabilities), (b) attention is given to the appropriate level of explanation of the climate variable, which operationally means that before one computes an interrater reliability, he/she should be reasonably assured that subjects were perceiving the same set of events, and (c) individuals on whom estimates of agreement are based are relatively similar in regard to personalistic variables that relate to cognitive information processing of climate perceptions. It is hypothesized, therefore, that individuals tend to agree at substantially higher levels than reported previously in the climate literature, given the conditions specified. If this hypothesis is confirmed, then the concept of an organizational climate, or perhaps "position climate", is alive and well, although we may wish to adopt a different descriptor than "organizational" to indicate aggregate psychological climate perceptions. Finally, a somewhat obvious but nevertheless important point requires mention. If the environments (positions) in a sample are indeed homogeneous, and if the within-environment variation on a climate variable is low, then it follows that the $\underline{r_{WG(\bar{X}_i)}}$s can be quite high but the mean climate scores will not relate highly, or perhaps even moderately, to other environmental variables (e.g., structural variables). This is, of course, due to the restriction of range on the mean climate scores, and most likely other environmental variables. Thus, the points raised here regarding the effects of restriction of range on interrater reliability estimates such as ICC(1) extend directly to relations between aggregate climate scores and other variables. On the other hand, restriction of range is not as serious for climate data as it may be for other variables inasmuch as climate data often serve an important diagnostic function, such as ascertaining whether individuals in an environment, or each of a set of environments, perceive their pay and benefit programs as fair and equitable.

References

Bartko, J. J. On various intraclass correlation reliability coefficients. Psychological Bulletin, 1976, 83, 762-765.

Campbell, J. P., Dunnette, M. D., Lawler, E. E., III, & Weick, K. E., Jr. Managerial behavior, performance, and effectiveness. New York: McGraw-Hill, 1970.

Cohen, J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 1960, 20, 37-46.

Cooper, M. An exact probability test for use with Likert-type scales. Educational and Psychological Measurement, 1976, 36, 647-655.

Ebel, R. L. Estimation of the reliability of ratings. Psychometrika, 1951, 16, 407-424.

Endler, N. S., & Magnusson, D. Toward an interactional psychology of personality. Psychological Bulletin, 1976, 83, 956-974.

Finn, R. H. A note on estimating the reliability of categorical data. Educational and Psychological Measurement, 1970, 30, 71-76.

Guion, R. M. A note on organizational climate. Organizational Behavior and Human Performance, 1973, 9, 120-125.

Hellriegel, D., & Slocum, J. W., Jr. Organizational climate: Measures, research, and contingencies. Academy of Management Journal, 1974, 17, 255-280.

Hsu, L. Agreement or disagreement of a set of Likert-type ratings. Educational and Psychological Measurement, 1979, 39, 291-295.

Insel, P. M., & Moos, R. H. Psychological environments: Expanding the scope of human ecology. American Psychologist, 1974, 29, 179-188.

James, L. R. Aggregation bias in estimates of perceptual agreement. Journal of Applied Psychology, 1982, 67, 219-229.

James. L. R., Gent, M. J., Hater, J. J., & Coray, K. E. Correlates of psychological influence: An illustration of the psychological climate approach

to work environment perceptions. Personnel Psychology, 1979, 32, 563-588.

James, L. R., Hater, J. J., Gent, M. J., & Bruni, J. R. Psychological climate: Implications from cognitive social learning theory and interactional psychology. Personnel Psychology, 1978, 31, 781-813.

James, L. R., Hater, J. J., & Jones, A. P. Perceptions of psychological influence: A cognitive information processing approach for explaining moderated relationships. Personnel Psychology, 1981, 34, 453-477.

James, L. R., & Jones, A. P. Organizational climate: A review of theory and research. Psychological Bulletin, 1974, 81, 1096-1112.

James, L. R., & Jones, A. P. Perceived job characteristics and job satisfaction: An examination of reciprocal causation. Personnel Psychology, 1980, 33, 97-135.

James, L. R., & Sells, S. B. Psychological climate: Theoretical perspectives and empirical research. In D. Magnusson (Ed.), Toward a psychology of situations: An interactional perspective. Hillsdale, NJ: Lawrence Erlbaum, 1981.

Jones, A. P., & James, L. R. Psychological climate: Dimensions and relationships of individual and aggregated work environment perceptions. Organizational Behavior and Human Performance, 1979, 23, 201-250.

Jones, E. E., & Gerard, H. B. Foundations of social psychology. New York: Wiley, 1967.

Joyce, W. F., & Slocum, J. W., Jr. Climates in organizations. In S. Kerr (Ed.), Organizational behavior. Columbus, Ohio: Grid, 1979.

Kim, J. S. Relationships of personality to perceptual and behavioral responses in stimulating and nonstimulating tasks. Academy of Management Journal, 1980, 23, 307-319.

Lord, F. M., & Novick, M. R. Statistical theories of mental test scores.

Reading, MA: Addison-Wesley, 1968.

Maxwell, S. E., Camp, C. J., & Arvey, R. D. Measures of strength of association: A comparative examination. Journal of Applied Psychology, 1981, 66, 525-534.

Mischel, W. Toward a cognitive social learning reconceptualization of personality. Psychological Review, 1973, 80, 252-283.

Naylor, J. D., Pritchard, R. D., & Ilgen, D. R. A theory of behavior in organizations. New York: Academic Press, 1980.

Newman, J. E. Understanding the organizational structure - job attitude relationship through perceptions of the work environment. Organizational Behavior and Human Performance, 1975, 14, 371-397.

Payne, R. L., Fineman, S., & Wall, T. D. Organizational climate and job satisfaction: A conceptual synthesis. Organizational Behavior and Human Performance, 1976, 16, 45-62.

Payne, R. L., & Pugh, D. S. Organizational structure and climate. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology. Chicago: Rand McNally, 1976.

Powell, G. N., & Butterfield, D. A. The case for subsystem climates in organizations. Academy of Management Review, 1978, 3, 151-157.

Roberts, K. H., Hulin, C. L., & Rousseau, D. M. Developing an interdisciplinary science of organizations. San Francisco: Jossey-Bass, 1978.

Schneider, B. Organizational climate: An essay. Personnel Psychology, 1975, 28, 447-479.

Schneider, B. Work climates: An interactionist perspective. In N. W. Feimer & E. S. Geller (Eds.), Environmental psychology: Directions and perspectives. New York: Praeger, in press.

Schneider, B., Parkington, J. J., & Buxton, V. M. Employee and customer perceptions of service in banks. Administrative Science Quarterly, 1980,

_25_, 252-267.

Selvage, R. Comments on the analysis of variance strategy for computation of intraclass reliability. _Educational and Psychological Measurement_, 1976, _36_, 605-609.

Shrout, P. E., & Fleiss, J. L. Intraclass correlations: Uses in assessing rater reliability. _Psychological Bulletin_, 1979, _86_, 420-428.

Stotland, E., & Canon, L. K. _Social psychology: A cognitive approach_. Philadelphia: Sanders, 1972.

Winer, B. J. _Statistical principles in experimental design_. New York: McGraw-Hill, 1971.

Woodman, R. W., & King, D. C. Organizational climate: Science or folklore? _Academy of Management Review_, 1978, _3_, 816-826.

Reference Notes

1. James, L. R., Wolf, G., & Demaree, R. G. Estimating interrater reliability in incomplete designs. Fort Worth, TX: Institute of Behavioral Research, Texas Christian University, 1981.

Footnotes

Support for this project was provided under Office of Naval Research Contract N00014-80-C-0315, Office of Naval Research Project NR170-904. Opinions expressed are those of the author and are not to be construed as necessarily reflecting the official view or endorsement of the Department of the Navy. The author wishes to thank David W. Bracken, Robert G. Demaree, Allan P. Jones, Benajmin Schneider, S. B. Sells, and Gerrit Wolf for their helpful suggestions and advice. Requests for reprints should be sent to Lawrence R. James, School of Psychology, Georgia Institute of Technology, Atlanta, Georgia 30332.

[1]Inasmuch as the $r_{WG(\bar{X}_i)}$ prodecure does not weight items explicitly, the items were not weighted explicitly (e.g., component scores) in the calculation of means. This provided as comparable a base as possible for contrasting reliability estimates provided by the *ICC*(1) and $r_{WG(\bar{X}_i)}$ approaches.

Table 1

Intraclass Correlation Based on Twenty Raters

in Each of Two Environments

Scale for Variable X	Frequencies of Scores in Environment 1	Frequencies of Scores in Environment 2
1	0	0
2	2	2
3	16	15
4	2	3
5	0	0
Mean:	3.00	3.05
Variance:	.211	.261

Analysis of Variance

Source	df	SS	MS	
Between-Environment	1	.025	.025	$F = .106^{NS}$
Within-Environment	38	8.959	.236	

Intraclass Correlation

$$\mathit{ICC}(1) = \frac{.025 - .236}{.025 + (19)(.236)}$$

$$= -.047$$

$$\cong .00$$

Note: NS = not significant at $p < .05$.

Table 2

Illustrations of Within-Group *ICC* and $\underline{r_{WG(\bar{X}_i)}}$ for

a Single Group of Raters

A. Data

Item	Rater 1	2	3	4	5	6	Mean	$s_{X_j}^2$
1	6	6	7	7	7	7	6.67	.27
2	7	6	6	7	6	6	6.3	.27
3	7	7	7	6	6	6	6.5	.30
4	6	7	6	7	6	7	6.5	.30
Mean	6.5	6.5	6.5	6.75	6.25	6.50		

B. Within-Group *ICC*

Source	df	SS	MS
Between-Item	3	.411	.137
Within-Item	20	5.70	.285
Between Rater	5	.50	.10
Residual	15	5.20	.347

ICC ≅ .00

C. $r_{WG(\bar{X}_i)}$

$$\overline{s_X^2} = (.27 + .27 + .30 + .30)/4 = .285$$

$$\sigma_E^2 = (7^2 - 1)/12 = 4.0$$

$$r_{WG(\bar{X}_i)} = \frac{4[1 - (.285/4.0)]}{4[1 - (.285/4.0)] + (.285/4.0)}$$

$$= .98$$

Table 3

Percentile Ranks, Variance Proportions, and Between-Group Differences

for Technician Sample on Individual Difference Variables

Variable	Percentile Rank of Technician Sample Mean[a]	Technician Sample Variance/ Norm Sample Variance[b]	η^2 Based on One-Way ANOVA[c]
Gordon Personal Inventory			
Cautiousness	86	.54	.04*
Original Thinking	82	.50	.04
Personal Relations	84	.58	.03*
Vigor	77	.48	.04*
Gordon Personal Profile			
Ascendency	67	.50	.03
Responsibility	95	.48	.03
Emotional Stability	79	.52	.04
Sociability	48	.45	.04
Bennett Test of Mechanical			
Comprehension--Form BB	75	.51	.05*

* $p < .05$ for F-ratio associated with eta-square (η^2).

[a] Total sample size for all offices (k=87) was 3,050 for the Bennett and an average of 2,770 for the personality inventory and profile. The norm samples used here for all variables were based on high school students.

[b] The total sample variances and mean-square errors from the one-way ANOVAs were essentially identical. Proportions are based on total sample variances. F-max tests on norm sample variance/technician sample variance were significant ($p < .05$) for all variables.

[c] Eta-square for one-way ANOVAs to test for mean differences among the 87 offices.

Table 4

Comparison of Between-Office and Within-Office Estimates

of Interrater Reliability on Eleven Climate Dimensions

Climate Dimension	Number of Items (1)	Coefficient Alpha (2)	ICC(1) (3)	Variance of Office Means[1] (4)	Average Within-Office Variance[2] (5)	$r_{WG(\overline{X}_i)}$ Range (6)	$r_{WG(\overline{X}_i)}$ Percent $\geq$.70 (7)
1. Adequacy of Benefits	7	.84	.08	.04	.42	.87 - .96	100%
2. Affirmative Action--Negative Consequences	3	.90	.04	.04	.82	.35 - .76	22
3. Involvement	10	.89	.03	.02	.45	.86 - .96	100
4. Affirmative Action--Positive Consequences	4	.74	.01	.01	.46	.63 - .89	95
5. Habitability	4	.83	.07	.04	.38	.52 - .92	95
6. Overload	4	.73	.05	.03	.51	.61 - .88	88
7. Resources	4	.79	.03	.03	.56	.69 - .93	98
8. Equity of Pay	4	.86	.08	.07	.64	.50 - .89	93
9. Fairness - Management	4	.76	.06	.03	.42	.66 - .91	98
10. Employment Security	9	.87	.10	.05	.23	.81 - .95	100
11. Opportunities for Advancement	3	.75	.03	.05	1.00	.39 - .91	65

Note. K = 60 offices, N = 7,180.

[1]Variance of K=60 mean climate scores for offices.

[2]Average within-office variance of technicians' climate composite scores.

FILMED

5-8